Cluster analysis of the organic peaks in bulk mass spectra obtained during the 2002 New England Air Quality Study with an Aerodyne aerosol mass spectrometer

C. Marcolli, M. R. Canagaratna, D. R. Worsnop, R. Bahreini, J. A. de Gouw, C. Warneke, P. D. Goldan, W. C. Kuster, E. J. Williams, B. M. Lerner, et al.

▶ **To cite this version:**

C. Marcolli, M. R. Canagaratna, D. R. Worsnop, R. Bahreini, J. A. de Gouw, et al.. Cluster analysis of the organic peaks in bulk mass spectra obtained during the 2002 New England Air Quality Study with an Aerodyne aerosol mass spectrometer. Atmospheric Chemistry and Physics Discussions, 2006, 6 (3), pp.4601-4641. hal-00301422

HAL Id: hal-00301422
https://hal.science/hal-00301422

Submitted on 12 Jun 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Atmos. Chem. Phys. Discuss., 6, 4601–4641, 2006
www.atmos-chem-phys-discuss.net/6/4601/2006/
© Author(s) 2006. This work is licensed
under a Creative Commons License.

Atmospheric Chemistry and Physics Discussions

Cluster analysis of the organic peaks in bulk mass spectra obtained during the 2002 New England Air Quality Study with an Aerodyne aerosol mass spectrometer

C. Marcolli[1], M. R. Canagaratna[2], D. R. Worsnop[2], R. Bahreini[3,4], J. A. de Gouw[3,4], C. Warneke[3,4], P. D. Goldan[3], W. C. Kuster[3], E. J. Williams[3,4], B. M. Lerner[3,4], J. M. Roberts[3], J. F. Meagher[3], F. C. Fehsenfeld[3], M. L. Marchewka[5], S. B. Bertman[5], and A. M. Middlebrook[3]

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Switzerland
[2]Aerodyne Research Incorporated, Billerica, MA, USA
[3]National Oceanic and Atmospheric Administration, Earth System Research Laboratory, Chemical Sciences Division, Boulder, CO, USA
[4]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA
[5]Department of Chemistry, Western Michigan University, Kalamazoo, MI, USA

Received: 3 April 2006 – Accepted: 9 April 2006 – Published: 12 June 2006

Correspondence to: C. Marcolli (claudia.marcolli@env.ethz.ch)

Abstract

We applied hierarchical cluster analysis to an Aerodyne aerosol mass spectrometer (AMS) bulk mass spectral dataset collected aboard the NOAA research vessel Ronald H. Brown during the 2002 New England Air Quality Study off the east coast of the United States. Emphasizing the organic peaks, the cluster analysis yielded a series of categories that are distinguishable with respect to their mass spectra and their occurrence as a function of time. The differences between the categories mainly arise from relative intensity changes rather than from the presence or absence of specific peaks. The most frequent category exhibits a strong signal at m/z 44 and represents oxidized organic matter most probably originating from both, anthropogenic as well as biogenic sources. On the basis of spectral and trace gas correlations, the second most common category with strong signals at m/z 29, 43, and 44 contains contributions from isoprene oxidation products. The third through the fifth most common categories have peak patterns characteristic of monoterpene oxidation products and were most frequently observed when air masses from monoterpene rich regions were sampled. Taken together, the second through the fifth most common categories represent as much as 5 μg/m^3 organic aerosol mass – 17% of the total organic mass – that can be attributed to biogenic sources. These numbers have to be viewed as lower limits since the most common category was attributed to anthropogenic sources for this calculation. The cluster analysis was also very effective in identifying a few contaminated mass spectra that were not removed during pre-processing. This study demonstrates that hierarchical clustering is a useful tool to analyze the complex patterns of the organic peaks in bulk aerosol mass spectra from a field study.

1 Introduction

Particles in the atmosphere play important roles in the Earth's climate and human health. Ambient measurements of their abundance, size, and chemical composition

ACPD

6, 4601–4641, 2006

lead to an increased understanding of their sources and how they evolve in the atmosphere. The Aerodyne aerosol mass spectrometer (AMS) has been designed to measure size-resolved mass distributions and total mass loadings of non-refractory chemical species from submicron particles (Jayne et al., 2000; Allan et al., 2003a; Jimenez et al., 2003). In the AMS, particles are focused into a vaporizer located within the electron impaction ionization source region of a quadrupole mass spectrometer. When it is operated in the mass spectrum mode, mass spectra from m/z 1–300 can be collected from several particles with good time resolution yet without particle size information. By extracting the mass spectral ion signals for various species, it has been shown to yield quantitative aerosol compositions for sulphate, nitrate, ammonium and total organic particulate matter (Allan et al., 2003b, 2004a; Bahreini et al., 2003; Schneider et al., 2004).

The AMS signals in most m/z channels are due to organic species; however, the exact assignment of peaks in complex mixtures such as those found in atmospheric aerosol particles is impossible. For organic substances, the AMS has been shown to generate mass spectra that are consistent with the electron impact ionization spectra found in standard mass spectrometry libraries yet with generally more fragmentation (Allan et al., 2003b; Alfarra, 2004). Furthermore, increased fragmentation has been observed in the AMS with increased vaporizer temperature (Alfarra, 2004). Organic functional groups can be identified from the small (m/z<50) ion fragments and information about the oxidation and saturation state of the organic aerosol fraction can be retrieved by means of the so-called ion series or delta analysis (McLafferty and Tureek, 1993; Allan et al., 2003b; Schneider et al., 2004; Bahreini et al., 2003, 2005; Alfarra et al., 2004; Canagaratna et al., 2004). However, specific patterns of the organic signals can not be captured by this method alone. Although discrimination between fragments with the same nominal m/z such as those listed in Table 1 is now possible with a high resolution time-of-flight AMS, several research groups have obtained low resolution data with the quadrupole AMS. The question therefore arises as to what other information about the aerosol organic material can be extracted from an AMS dataset.

One method that has been used involves multivariate linear regressions, separating the ambient mass spectra into linear combinations of two or three components based on m/z 44, m/z 43, and m/z 57 (Zhang et al., 2005a). Scatter plots of organic peak intensities have also been used to compare ambient AMS data with reference spectra (Alfarra et al., 2006[1]).

A hierarchical clustering algorithm has been successfully applied to categorize mass spectra from individual particles measured by the Particle Analysis by Laser Mass Spectrometry (PALMS) instrument (Murphy et al., 2003). In the present study, we apply this algorithm to a quadrupole AMS dataset of particle ensembles collected during the 2002 New England Air Quality Study (NEAQS) to observe the occurrence of typical organic mass spectral features.

2 Experimental methods

The AMS dataset was collected aboard the NOAA research vessel Ronald H. Brown off the east coast of the United States between 18 July and 10 August 2002 as part of NEAQS 2002. A map showing the ship track is shown in Fig. 1. For this study, many different types of air masses and sources were sampled (e.g., urban from Boston and New York City and biogenic from Maine) and therefore a detailed cluster analysis was performed to see whether or not specific mass spectral signatures of the several different types of general sources in the bulk mass spectra could be detected. Mass spectra from m/z 1–300 and particle time-of-flight mass distributions were recorded every 2 min with 1 min sampling times in each mode for most of the study. More experimental details of the AMS dataset and how the mass loadings were measured

[1] Alfarra, M. R., Paulsen, D., Gysel, M., Garforth, A. A., Dommen, J., Prevot, A. S. H., Worsnop, D. R., Baltensperger, U., and Coe, H.: A mass spectrometric study of secondary organic aerosols formed from the photooxidation of anthropogenic and biogenic precursors in a reaction chamber, Atmos. Chem. Phys., submitted, 2006.

will be reported elsewhere (Middlebrook et al., 2006[2]). Volatile organic compounds (VOCs) were measured by on-line gas chromatography-mass spectrometry (Goldan et al., 2004) and proton-transfer-reaction mass spectrometry (de Gouw et al., 2003), and ozone measurements were obtained using a commercial UV absorbance instrument (Thermo Environmental Instruments Inc., Model 49C).

The AMS non-refractory mass loadings for organic material, sulphate, ammonium, and nitrate are shown in Fig. 2 along with the wind direction during sampling. As previously reported, the overall submicron aerosol composition during this study was predominantly organic and sulphate (Quinn and Bates, 2003; Bates et al., 2005). The highest loadings of both sulphate and organic mass occurred when the wind direction was from the southwest where the air masses had recently come from urban sources. The non-refractory aerosol mass loadings were significantly lower and dominated by organic material when the winds were from the north where biogenic emissions were prevalent.

In this manuscript we will make use of the photochemical age of an airmass, an isoprene source term, and a delta analysis of the AMS mass spectra. The photochemical age of the air masses was determined from the VOCs by examining the ratios of toluene to benzene, and the amount of isoprene from the source was extrapolated from the ratio of methyl vinyl ketone plus methacrolein to isoprene (see Eqs. 1, 2a, and 2b in de Gouw et al., 2005). For the delta analysis, each m/z in the spectrum corresponds to a delta value equal to m/z–14n+1 where n is an integer (McLafferty and Tureek, 1993). The delta value of a particular m/z provides an indication of the functionality, given that hydrocarbons typically differ by a CH_2 group which is 14 amu, with negative delta values generally indicating unsaturated hydrocarbons and positive delta values generally indicating oxidized hydrocarbons. The intensity-weighted averages of the

[2]Middlebrook, A. M., Matthew, B. M., Worsnop, D. R., Canagaratna, M. R., Bates, T. S., Quinn, P. K., Kuster, W. C., Goldan, P. D., Meagher, J. F., and Fehsenfeld F. C.: Evidence of organic nitrate in polluted aerosols during the NEAQS 2002 Field Study, Geophys. Res. Lett., in preparation, 2006.

delta values are then calculated for various ranges of carbon number (Bahreini et al., 2005).

The hierarchical clustering algorithm that was developed for PALMS (Particle Analysis by Laser Mass Spectrometry) data (Murphy et al., 2003) is briefly described here. The most similar spectra or clusters are sequentially combined until a stopping condition is met. The algorithm uses the dot product between mass spectra as the similarity criterion: the most similar spectra or clusters are assigned the same category and their mass spectra are averaged. Mathematically, the total signal in each spectrum is normalized to unit intensity and then each spectrum is represented as a vector:

$$\boldsymbol{A} = \left[a_1, a_2, a_3, \text{etc.}\right] \quad (1)$$

where a_1, a_2, a_3, etc. are the normalized peak intensities at various m/z. The dot product of two normalized vectors, $\boldsymbol{A}$ and $\boldsymbol{B}$, is simply:

$$\boldsymbol{A} \times \boldsymbol{B} = a_1 \times b_1 + a_2 \times b_2 + a_3 \times b_3 + \text{etc.} \quad (2)$$

This dot product is the cosine of the angle between the two vectors: if it is zero the spectra are orthogonal and if it is one the spectra are identical. In hierarchical clustering, the data are searched for the pair of spectra with the highest dot product. Those two spectra are averaged and renormalized to produce a category that replaces them. Then the dataset, now smaller by one spectrum, is again searched for the highest dot product, whether it is between two spectra or a spectrum and a category. These are averaged and the process is continued until all the remaining dot products are less than a stopping condition (strict criterion). For categories containing a small number of spectra, the stopping condition is loosened (loose criterion) to allow less-closely matched spectra to be combined in small categories.

The PALMS hierarchical clustering algorithm was adapted to the AMS mass spectra. Spectra obtained during the field study from ammonium nitrate calibration particles or that were known to contain contaminants were initially removed, resulting in 12 730 spectra for the cluster analysis. The peaks known to be dominated by air or by inorganic

ions (m/z 14–24, 28, 32–34, 36, 39, 40, 47–50, 64–66, 80–83, 98–100) were excluded and the peaks at m/z 29, 30, 38, and 44 were corrected for contributions from air. The average noise levels as determined by the method described by Allan et al. (2003a) have been subtracted from the peak signals at each m/z and any negative values were set to zero. Finally, the total signal of each spectrum was normalized to unit intensity. In the application for the PALMS datasets, a combination of the linear scaled peak intensities and the log scaled raw spectrum was used. For the present work, only the linear peak intensities were used. Using the log of the signals intensified the noise level of the high m/z peaks and did not contain useful information since most of the spectral signal was in the smaller ion fragments. The diversity (or number of categories) and the relative occurrence of the categories for approximately every two hours were calculated by forming groups of sixty of the 2 min averaged spectra as a function of time.

3 Results and discussion

3.1 Categorization

The AMS mass spectra in this study were from ambient atmospheric aerosols, which are composed of numerous organic compounds with varying concentrations rather than few individual organic species. Thus, these data did not form distinct clusters per se, and the choice of the stopping condition and the number of resulting categories was arbitrary. For our dataset, the 12 730 spectra were first divided into four subsets by time. To reduce computing time, the clustering algorithm was run separately for each subset until the number of categories was low enough to be easily handled in one set. The stopping condition for the last pass was optimized to yield clusters with different spectral features as well as distinct time sequences. The criteria for each pass and the resulting number of categories are shown in Table 2. After the categories were established, a final pass of all the spectra was performed to ensure that they were placed in the appropriate categories. We stopped with 48 categories – the largest

one comprising 9505 of the total of 12 730 spectra, the smallest ones containing one spectrum each.

Table 3 lists the most prominent peaks in the top 15 categories and the averaged mass spectra of some important categories are shown in Fig. 3. The top 5 categories comprise over 92% of the spectra. The categories mainly differ in the relative intensity rather than in the presence or absence of specific peaks. Most of them are dominated by peaks with m/z below 50. Thus, the sampled organic aerosol fraction either consisted of quite small molecules or strongly fragmented in the mass spectrometer. The Aerodyne mass spectra collected during the MINOS campaign by Schneider et al. (2004) also show a strong dominance of low m/z peaks whereas around 40% of the total organic signal during the ACE-Asia campaign was due to fragments with m/z above 50 (Bahreini et al., 2003). The individual spectra are assigned to the category which gives the best match based on the dot product as the similarity criterion. While still meeting the criterion, the sequence of most intense peaks for individual category members can deviate from the one of the averaged spectrum: 8% of category 1 spectra exhibit the strongest signal at m/z 43 instead of m/z 44 as the average in Fig. 3 shows; in category 2, 46% of the spectra show m/z 29 as the strongest peak, 31% m/z 43, and 19% m/z 44, reflecting the similar intensities of m/z 29, 43, and 44 in the averaged spectrum.

Some of the most abundant peaks can be unambiguously assigned to specific organic fragments as shown in Table 1. Mass m/z 44 is a sign of oxidized organic species since it is due to CO_2^+ e.g. from oxo- and dicarboxylic acids as well as highly oxygenated species (Alfarra, 2004). It is strong in category 1 (Fig. 3), thus, this category can be considered as highly oxidized. Mass m/z 30 was included in this analysis since it could be due to NO^+ from organic nitrates, nitro aromatics, and possibly other organo-nitrogen or organo-oxygen containing species in addition to inorganic nitrates. It is the most abundant peak in category 7 (Fig. 3), the second most abundant peak in category 13, and is common in many other categories (Table 3). The peaks at m/z 69, 119, and 169 have been observed in fluorinated hydrocarbon oil from mechanical pumps (J.

ACPD
6, 4601–4641, 2006

EGU

Jimenez and P. Silva, personal communication). Mass m/z 69 and 169 are the second most and the most abundant ones in categories 10 (with 60 spectra) and 14 (with 24 spectra), respectively (Fig. 3).

In many of the categories, m/z 57 is typically not a major peak. A strong peak at m/z 57 along with m/z 55 and other hydrocarbon fragments differing by 14 amu have been found in the AMS spectra from direct emissions of diesel exhaust (Canagaratna et al., 2004) and this distinct pattern has been observed in urban mass spectra during the morning (Alfarra et al., 2004; Allan et al., 2003b; Zhang et al., 2005a, b). During the NEAQS 2002 study, however, the only time when a relatively high m/z 57 was observed was when the ship's exhaust was inadvertently sampled while changing position, and these spectra were removed from the cluster analysis during pre-processing. The lack of fresh organic aerosol spectra during this study is consistent with the VOC data which nearly always exhibited some degree of aging (de Gouw et al., 2005).

3.2 Comparison with mass spectra

Very recently, mass spectra of secondary organic aerosols formed from the oxidation of anthropogenic and biogenic precursors have been measured by different research groups with the quadrupole AMS (Alfarra et al., 2006[1]; Bahreini et al., 2005; Zhang et al., 2005a; Allan et al., 2004b). Secondary organic aerosol has also recently been observed in chamber studies of isoprene oxidation (Kroll et al., 2005), and AMS mass spectra of these particles have been obtained in similar experiments (unpublished data). The laboratory spectra are shown in Fig. 4 together with reference spectra for fulvic acid (Alfarra et al., 2006[1]) and oxidized urban aerosol (Pittsburgh OOA) (Zhang et al., 2005a). These spectra are compared to the top 5 categories to find similarities and their correlations are shown in Fig. 5. Since aerosols from different sources may have been sampled at the same time by the AMS spectrometer during the field study, the measured mass spectra might be dominated by one type of aerosol or contain a mixture of aerosol types. Because of its abundance throughout the whole campaign, category 1 spectral features are the most likely to be mixed into other categories re-

sulting in a relatively large m/z 44 signal. Furthermore, other peaks in the spectra may form distinct patterns in the absence of m/z 44. Therefore, the correlations were carried out with and without m/z 44. The complete set of peaks used for these correlations are at m/z 12, 13, 25, 26, 27, 29, 31, 37, 38, 41, 42, 43, 44, 45, 51, 53, 55, 57, 67, 69, 71, 73, 85, 87, 89, and 91.

The best correlations of category 1 occur with the spectrum of oxidized organic aerosols obtained in Pittsburgh, which become distinctly better than the ones with meta-xylene when m/z 44 is excluded (Fig. 5a). Therefore, category 1 can be considered as highly oxidized. The patterns observed in categories 2–5 are quite similar to spectra obtained from oxidation of both anthropogenic and biogenic precursors (Alfarra et al., 2006[1]; Bahreini et al., 2005) and these particles are less oxidized than when m/z 44 dominates the spectrum (Alfarra, 2004). The category 2 spectrum has a large peak at m/z 29, similar signal intensities at m/z 43, and m/z 44 and relatively low signal at m/z 41, m/z 55, and m/z 57 which signifies the prevalence of carbonyls (C=O) from either ketones, aldehydes, and carboxylic acids. This category shows the highest correlation with the Pittsburgh oxidized organic aerosol as well as the isoprene oxidation spectra (Fig. 5b) and could be formed by a linear combination of both. Category 3 most closely resembles the spectrum observed of aerosols produced by α-pinene oxidation (Alfarra et al., 2006[1]), especially when m/z 44 is excluded (Fig. 5c). Hence, it could be characteristic of more strongly oxidized products than present in the reference. Category 4 is extremely similar to the spectra formed by oxidation of α-pinene or 1,3,5-trimethylbenzene (TMB) (Alfarra et al., 2006[1]) (Fig. 5d) and the pattern of the main peaks of category 5 has the highest resemblance to the one formed during α-pinene ozonolysis experiments (Bahreini et al., 2005) (Fig. 5e).

In addition to subtle differences in the mass spectra of the low m/z peaks from various precursors, the high m/z (C5–C6 and C7–C15) fragments have distinct delta patterns. The biogenic precursors generate spectra with an average negative delta value whereas the spectra from cycloalkene precursors and m-xylene have an average positive delta value (Bahreini et al., 2005). The average delta patterns for the top 5

categories (Fig. 6) are similar to each other and more similar to those from biogenic precursors than those from anthropogenic precursors, implying that the organic mass of higher m/z fragments in these categories is biogenic in origin. Note that 23% of the total organic signal is represented by these peaks. Indeed all of the top 15 categories, with the exception of category 14, have comparable average delta patterns. The delta pattern of category 5 is nearly identical to that from α-pinene ozonolysis products (Bahreini et al., 2005) which strongly indicates that secondary organic aerosol from α-pinene ozonolysis was present in these spectra. However it must be noted that the delta patterns for the aerosol products of only one aromatic anthropogenic precursor compound (meta-xylene) have been investigated (Bahreini et al., 2005).

3.3 Trends with time and wind direction

To judge whether the categories are related to certain time periods, we investigated them as a function of time (Fig. 7). The top two panels show the organic mass loadings and the diversity of the organic aerosol fraction over the whole measurement period. It can be seen that the diversity – i.e. the number of different categories within groups of 60 consecutive measurements – is highest at times when the overall mass loading is lowest. Since category 1 comprises 75% of the spectra, there is a very strong anti-correlation between the diversity and this category. The low diversity for high mass loadings may be an indication of the organic components repartitioning among the particles as they age (Marcolli et al., 2004). However, the high diversity at low mass loadings might be to some extent a statistical effect, since for averaging times of 1 min and at low mass loadings only few particles are responsible for the detected signal at a specific m/z channel (Bahreini et al., 2003).

Category 1 was dominant during almost the whole measurement period, being most abundant at times when the organic mass loading was high. Its incidence dropped notably during three periods: on 23 July, 3 August, and 6 August (Fig. 7). In all three instances, the organic mass loadings were much lower, due to either the passage of cold fronts or major wind direction shifts from the southeast to the southwest or north in

ACPD
6, 4601–4641, 2006

a westerly fashion (Fig. 2). Although the contribution to any wind direction is dominated by category 1 spectra, the relative contribution for category 1 spectra is the highest when the winds are from the east or south (Fig. 8). Category 2 is prevalent on 23–24 July, 3 August, and 5–6 August during wind shifts at the end of the high organic mass events and is highest on 5 August when categories 3 through 5 are altogether absent. During these time periods the winds are from the southwest (Figs. 2 and 8) and have passed over the Boston urban area as well as an isoprene "hot spot" (Fig. 1a). This further supports the notion of category 2 as being indicative of isoprene oxidation products mixed together with aged urban (Pittsburgh OOA) aerosol. Categories 3 through 5 are most common mainly during two time periods (24–26 July and 6–7 August) when the organic mass loading was relatively low and winds are from the north (Figs. 2 and 8). Because northerly winds come from over Canada with relatively low isoprene and high monoterpene emissions (Fig. 1), the incidence of categories 3 through 5 supports the interpretation of these categories as originating from monoterpene oxidation products. Although categories 7 and 13 both have relatively large peaks at m/z 30, their time trends are different: Category 7 was observed mainly on three days – 21 July, 3 August and 7 August – whereas category 13 was observed mainly on 7 August, the day with the lowest submicron aerosol mass throughout the whole campaign. Categories 10 and 14 occurred mainly on 10 August between 03:00 and 16:00 UTC, when the ship was in Chesapeake Bay. The ship's movement during this time did not create sufficient relative wind off the bow to sample air away from the ship. The clear signature of fluorocarbons in categories 10 and 14 indicates that the AMS sampled mechanical pump exhaust from other sampling equipment on the ship.

Other top 15 categories (6, 8, 9, and 11, not shown in Figs. 3 or 7) were observed most frequently during low mass periods when the spectra showed a large variation of composition as is reflected by the high diversity (Fig. 7b). They contain between 53 and 259 spectra and exhibit the main peak at m/z 27 (category 8), m/z 29 (category 9), and m/z 43 (category 11), indicating a low degree of oxidation at these low mass periods. The exception is category 6 which has a large peak at m/z 44. Category 11 with hardly

any signal at m/z 44 was only present when the organic mass was below 4 μg/m^3.

3.4 Diurnal and photochemical age trends

The diurnal cycles of the total organic mass loading averaged over the whole sampling period and the shortwave radiation as well as the average relative occurrence of the top 5 categories are shown in Fig. 9. Although spectra were averaged that were collected at different locations and under varying weather conditions, it still allows for summarizing general trends: The particulate organic matter exhibits the highest values from 14:00 to 18:00 UTC (10:00 to 14:00 local time) when the solar radiation was at a maximum during the afternoon. This is consistent with secondary organic aerosol formation by gas-to-particle conversion that contributed to the total particulate organic mass. Category 1 exhibits a similar diurnal cycle, but with a wider maximum. Therefore, gas-to-particle conversion may only partly explain the diurnal cycle of category 1, indicating that further processing might take place within the particles. Category 2 has a significant drop in frequency during midday (14:00 to 19:00 UTC or 10:00 to 15:00 local time), whereas the frequency of category 1 spectra increases during this time period. Category 5 is not abundant at high solar irradiance and typically peaks in the morning.

Figure 10 shows the relative occurrence of the top five categories versus the photochemical age of their air masses. The fraction of spectra placed in category 1 increased with age and there is a decreasing trend for category 2 spectra. Since category 1 has a significant peak at m/z 44, this could indicate processing of the organic material detected in category 2 spectra to form more oxidized organic material. Alternatively, the increase in relative occurrence of category 1 spectra may be due to condensation of oxidized material from a different source. It must be noted that the sampled air masses probably originated from or passed over different source regions (Fig. 8), so the relative contribution of each category as a function of photochemical age is not necessarily indicative of changes to the aerosol composition as urban air ages. Nevertheless, the

ACPD
6, 4601–4641, 2006

relative peak intensity at m/z 44 increased over time in chamber studies of secondary organic aerosol formation (Alfarra et al., 2006[1]), which is indicative of further oxidation during aging and consistent with the field data. The distributions shown in Fig. 10 for categories 3–5 are broad, yet tend to be relatively more important in younger air masses. Since the photochemical age is determined from anthropogenic VOCs, the weak correspondence with the degree of oxidation observed in the mass spectra is consistent with the particulate organic matter in categories 3–5 being comprised of minor contributions from anthropogenic precursors that have been oxidized during transport and aging.

3.5 Correlations of categories 1–5 with gas phase species

The occurrence of categories 1 through 5 is shown as a function of time in Fig. 11 for 25–26 July, the time period when winds were from the north and the ship was sailing off the coast of Maine. Also shown are the organic mass loadings as well as the concentrations of ozone and selected volatile organic compounds (VOCs): isopropyl nitrate, isoprene from the source, and the sum of all the monoterpenes. As discussed by de Gouw et al. (2005) and shown in the top panel of Fig. 11, the organic mass during this study is correlated with a secondary anthropogenic species (isopropyl nitrate), despite the relatively high concentrations of isoprene and monoterpenes. For this time period, category 1 exhibits a distinct correlation with ozone. Indeed, 93% of the spectra when ozone concentrations are above 75 ppbv are placed in category 1 and 6% are placed in category 2. This is consistent with the relative contribution of m/z 44 to the total organic mass correlating with the fraction of calculated mass corresponding to secondary anthropogenic processes (de Gouw et al., 2005). Although the concentrations of monoterpenes and isoprene (not shown) declined during the morning on 25 July (10:00 to 14:00 UTC or 6:00 to 12:00 local time), the isoprene source concentrations remained high for several hours, characteristic of an increased loss rate of the monoterpenes and isoprene from photochemistry as the air passed over the water to the ship rather than decreasing emissions or a change of air mass. Category 4 and to

a lesser degree categories 3 and 5 correlate with the monoterpenes, which is consistent with their spectral similarities with α-pinene oxidation products as shown in Fig. 5. Whereas the rise of category 3 is simultaneous with the rise of the monoterpenes, its decline is delayed, which might be indicative of category 3 consisting of more oxidized, longer-lived monoterpene oxidation products. Category 2 correlates better with the isoprene source concentrations, corroborating its characterization as isoprene oxidation products. Note that the isoprene source term is determined from the first generation oxidation products (methacrolein and methyl vinyl ketone). Including the second generation products as well as the first generation monoterpene products (such as pinonaldehyde) might show other trends. Considering the quite constant wind direction on 25 July, the increase of the organic mass and the rise of category 1 are rather due to enhanced gas-to-particle conversion with the onset of solar illumination than with a change of air mass source region, implying that during this time period also the particulate organic mass represented by category 1 could have biogenic sources. The correlations between the various biogenic gas phase species and categories 2–5 are strong during the time period depicted in Fig. 11 when the anthropogenic VOC concentrations were relatively low. These trends (especially those with isoprene) are not as clear in other air masses due to the increased abundance of category 1 spectra as well as anthropogenic VOCs.

The findings shown here indicate that some of the organic material originated from biogenic species that may have been oxidized in the presence of anthropogenic pollution via either ozone, hydroxyl radicals, or nitrate radicals in the presence of nitrogen oxides. Categories 2–5 appear to be characteristic of biogenic secondary organic aerosols (Figs. 5 and 6) and their occurrence is correlated with biogenic gas phase species (Fig. 11). Categories 2–5 are not as oxidized as category 1 spectra (Fig. 3) and they decrease in occurrence during the middle of the day when category 1 was prominent (Fig. 9). It is possible that the occurrence of category 2–5 is greater in photochemically young night time air masses because oxidation processes in daytime air masses convert category 2–5 aerosol into oxidized aerosol with spectra that might

fall into category 1. The lack of correlation for category 1 with biogenic species does not preclude its organic material from being biogenic. Indeed, during 23 July from 15:30 to 18:00 UTC (11:30 to 14:00 local time) when 60–70% of the ozone is attributed to biogenic hydrocarbons as determined from the peroxycarboxylic nitric anhydrides (Williams et al., 1997; Marchewka et al., 2006[3]), 75% of the AMS spectra are placed in category 1 and 25% are placed in category 2. On the basis of the evidence gathered in this study – comparison with reference spectra, delta analysis, incidence as a function of wind direction, and correlation with gas phase species – category 1 cannot be conclusively assigned to either biogenic or anthropogenic particulate organic matter. Rather, it might contain anthropogenic and biogenic contributions that are not distinguished with the stopping condition used in the cluster analysis. Aerosol processing leading to similar oxidation products for anthropogenic and biogenic precursors or strong fragmentation in the mass spectrometer might be reasons for high spectral similarities between anthropogenic and biogenic oxidation products. A high biogenic contribution to category 1 would be in accordance with radiocarbon measurements which show that modern carbon (rather than carbon from fossil fuel) is prevalent in atmospheric aerosols in a variety of urban areas in the U.S. (Klinedienst and Currie, 1999; Lemire et al., 2002; Lewis et al., 2004; E. Edgerton, personal communication) and around the world (Shibata et al., 2004; Szidat et al., 2004).

3.6 Calculated biogenic organic mass from categories 2–5

Assuming all of the organic mass in categories 2–5 is due to biogenic species, the amount of biogenic organic mass can be calculated by an hourly average of the mass in those categories. This was done in two ways: 1) the hourly average organic mass was multiplied by the hourly occurrence of spectra in categories 2–5 and 2) on an

[3]Marchewka, M., Bertman, S. B., Roberts, J. M., Warneke, C., de Gouw, J., Kuster, W., Goldan, P., Williams, E., Lerner, B. M., Murphy, P., and Fehsenfeld, F. C.: Measurements of PANs during the New England Air Quality Study 2002, J. Geophys. Res., submitted, 2006.

hourly basis the total organic mass from spectra in categories 2–5 was divided by the total organic mass in all categories (the fraction of organic mass in categories 2–5) and multiplied by the average organic mass. Both calculations yielded similar results, indicating that the fraction of mass in categories 2–5 was nearly identical to the fraction of spectra in those categories. The calculated biogenic mass from the spectra is shown in Fig. 12 along with the biogenic mass determined from the fraction of organic mass attributed to biogenic processes based on the isoprene source concentrations (de Gouw et al., 2005). The agreement in the calculated biogenic mass from the two independent methods is quite good for the second part of the study and during high mass loading periods. During the time period off the coast of Maine, the biogenic mass from the categories is significantly higher than that from the gas phase parameterization likely because monoterpenes were not included. Because category 1 is the most abundant category during the campaign, an accurate quantification of the contribution of biogenic species to the organic aerosol mass strongly depends on the estimated biogenic contribution to category 1. If it was significant, the biogenic mass loadings shown in Fig. 12 would be a lower limit. With this in mind, the contribution of biogenic species to the total organic mass by categories 2–5 alone averaged 17% for the entire study and 10% (ranging from 0–50%) when the total organic mass was more than 10 μg/m^3. These values are comparable to 12% of the organic mass being attributed to biogenic sources for the whole dataset by the gas phase method (de Gouw et al., 2005).

3.7 Categories with a large peak at m/z 30

Both categories 7 and 13 contain m/z 30: in category 7 it is the dominant peak whereas it is the second most common peak after m/z 44 in category 13 (Fig. 3). Categories 7 and 13 do not clearly correlate with isopropyl nitrate. Category 7 is mainly observed between 03:00 and 17:00 UTC (23:00 and 13:00 local time), with a significant peak between 10:00 and 12:00 UTC (top panel of Fig. 13). At this time (06:00 to 08:00 local), NO_3 and N_2O_5 are depleted, HNO_3 is at a morning minimum, OH is beginning

to increase, and there are large concentrations of NO_x which can form HNO_3 and organic nitrates (Brown et al., 2004). In contrast, the category 13 spectra are observed throughout the day (top panel of Fig. 13). Although there are less data points to consider, the category 7 spectra have a strong decreasing trend with photochemical age (bottom panel of Fig. 13) compared to the spectra in categories 1–5 (Fig. 10). Furthermore, category 7 spectra represent over 30% of the freshest air masses sampled (ages<4 h). There are only 5 spectra from category 13 with corresponding gas phase data for calculating their photochemical ages; yet they are present in older air masses than the spectra in category 7. Thus, the relative contributions of aerosol "nitrate" at m/z 30 and oxidized organic material at m/z 44 in these categories appear to be attributable to differences in photochemical ages of the air masses sampled, with the older air masses (category 13 as well as all the top 5 categories) containing more oxidized organic material than the younger air masses (category 7).

In addition to being in categories 7 and 13, m/z 30 is present at lower levels in the other top 15 categories (Table 3). The ratios of m/z 30 to m/z 46 for the NEAQS 2002 data are typically greater than 5, which is larger than the ratio from laboratory calibrations of ammonium nitrate particles (around 2). Ratios larger than 4 for these peaks have also been observed in oxidation studies where ammonium nitrate was not present (Alfarra et al., 2006[1]; Bahreini et al., 2005). Other pieces of evidence indicate that ammonium nitrate is not prevalent during this study (Quinn and Bates, 2003; Brown et al., 2004; Bates et al., 2005; Middlebrook et al., 2006[2]). Therefore, m/z 30 is likely due to organic nitrate (if relatively more NO^+ than NO_2^+ is formed compared to ammonium nitrate), an oxygenated organic fragment (CH_2O^+), and/or an amine fragment (CH_4N^+). Unless a high resolution mass spectrometer is used to separate these isomers at m/z 30, the identification of the species contributing to this peak remains ambiguous.

ACPD
6, 4601–4641, 2006

3.8 Comparison of hierarchical clustering with other data analysis methods

Many mass spectra in the analyzed AMS dataset were very uniform with a strong peak at m/z 44 as the main feature. The remaining spectra exhibited considerable variation of spectral characteristics that seemed to be specific for different sources or aging processes. In a principal component analysis, such minor numbers of mass spectra could easily be missed, since this method is designed to capture the main variances of a dataset. Moreover, a principal component analysis per se is not able to extract characteristic spectra but produces factors that are not directly comparable with reference spectra. An alternative way of analyzing AMS data is their deconvolution based on mass spectral tracers as demonstrated by Zhang et al. (2005a). This method must anticipate the relevant peaks and might become inconvenient if a high number of tracers were required. In contrast to this technique, the cluster analysis neither needs a preselection of relevant peaks nor any additional information such as gas phase concentrations of correlated species or wind directions. The use of the dot product as a similarity criterion for clustering implies that the most intense peaks primarily determine the categorization and are therefore the common features of all spectra in a category. In the investigated dataset, peaks with m/z below 50 dominate the organic mass signals. Since they are fragments of larger molecules and indicative of functional groups, they cannot be used to identify the presence of specific organic molecules. Nevertheless, clustering based on these low m/z peaks was able to extract categories that match well the mass spectral signatures of monoterpene and isoprene oxidation products and identify spectra with contaminants. A prior knowledge of possible reference spectra might help to optimize the stopping condition and thus increase the relevance and specificity of the categories. Alternatively, a dataset could be clustered around a database of references allowing a direct assignment of at least a part of the spectra. Application of hierarchical cluster analysis to further AMS datasets and comparison with other methods is needed to evaluate the potential of this technique more thoroughly.

4 Conclusions

Hierarchical cluster analysis was for the first time applied to an AMS dataset. Exclusion of peaks with inorganic contributions except for the nitrate peaks allowed us to focus on the organic aerosol fraction. The cluster analysis yielded a series of categories that are distinguishable with respect to their mass spectra and their occurrence as a function of time and wind direction. Furthermore, the spectral signatures of low m/z peaks in the most common categories are similar to reference spectra from field campaigns or generated in laboratory chamber studies. On the basis of the combined information (spectral and trace gas correlations), category 2 contains contributions from isoprene oxidation products and the organic mass in categories 3 through 5 appears to be formed by oxidation of monoterpenes. Taken together, categories 2 through 5 represent as much as 5 μg/m^3 of biogenic organic aerosol mass, averaging 17% of the total organic mass for the entire study and only 10% of the total organic mass when the total organic mass is greater than 10 μg/m^3. In contrast, the source of the organic mass in the category 1 spectra is unclear, since the spectra were placed in that category because they had a large signal at m/z 44, which is indicative of highly oxidized organic material irrespective of its origin.

Acknowledgements. This work was supported by the Swiss National Foundation under contract No. 200020-10365/1 and the NOAA Health of the Atmosphere program. We thank T. Peter for making this collaboration possible and the crew of the NOAA research vessel Ronald H. Brown for their assistance during the 2002 NEAQS study. The aerosol group at NOAA PMEL headed by T. Bates and T. Quinn provided the inlet and helpful discussions about the data. Several people provided fruitful discussions about the data acquisition and analysis: M. Aldener, C. Brock, S. Brown, H. Coe, P. DeCarlo, E. Edgerton, C. Eubank, K. Hayden, P. Hudson, J. Jayne, J. Jimenez, J. Kimmel, B. Matthew, P. Murphy, C. Simons, H. Stark, D. Sueper, Q. Zhang, and P. Ziemann. We appreciate receiving unpublished AMS reference spectra from R. Alfarra and S. Murphy. We are especially grateful to J. Allan for developing the AMS data analysis software, M. Warshawsky for developing the original PALMS clustering algorithm, and D. Murphy for assistance with adapting it for the AMS data.

References

Alfarra, M. R., Coe, H., Allan, J. D., Bower, K. N., Boudries, H., Canagaratna, M. R., Jimenez, J. L., Jayne, J. T., Garforth, A. A., Li, S., and Worsnop, D. R.: Characterization of urban and rural organic particulate in the Lower Fraser Valley using two Aerodyne aerosol mass spectrometers, Atmos. Environ., (Pacific 2001 special issue), 38, 5745–5758, 2004.

Alfarra, M. R.: Insights into atmospheric organic aerosols using an aerosol mass spectrometer, PhD thesis under Prof. H. Coe, Univ. of Manchester, 2004.

Allan J. D., Jimenez, J. L., Williams, P. I., Alfarra, M. R., Bower, K. N., Jayne, J. T., Coe, H., and Worsnop, D. R.: Quantitative sampling using an Aerodyne aerosol mass spectrometer, 1. Techniques of data interpretation and error analysis, J. Geophys. Res., 108(D3), 4090, doi:10.1029/2002JD002358, 2003a.

Allan J. D., Alfarra, M. R., Bower, K. N., Williams, P. I., Gallagher, M. W., Jimenez, J. L., McDonald, A. G., Nemitz, E., Canagaratna, M. R., Jayne, J. T., Coe, H., and Worsnop, D. R.: Quantitative sampling using an Aerodyne aerosol mass spectrometer, 2. Measurements of fine particulate chemical composition in two UK cities, J. Geophys. Res., 108(D3), 4091, doi:10.1029/2002JD002359, 2003b.

Allan, J. D., Delia, A. E., Coe, H., Bower, K. N., Alfarra, M. R., Jimenez, J. L., Middlebrook, A. M., Drewnick, F., Onasch, T. B., Canagaratna, M. R., Jayne, J. T., and Worsnop, D. R.: A generalised method for the extraction of chemically resolved mass spectra from Aerodyne aerosol mass spectrometer data, J. Aerosol Sci., 35(7), 909–922, doi:10.1016/j.jaerosci.2004.02.007, 2004a.

Allan, J. D., Bower, K. N., Coe, H., Boudries, H., Jayne, J. T., Canagaratna, M. R., Millet, D. B., Goldstein, A. H., Quinn, P. K., Weber, R. J., and Worsnop, D. R.: Submicron aerosol composition at Trinidad Head, California, during ITCT 2K2: Its relationship with gas phase volatile organic carbon and assessment of instrument performance, J. Geophys. Res., 109, D23S24, doi:10.1029/2003JD004208, 2004b.

Bahreini, R., Jimenez, J. L., Wang, J., Flagan, R. C., Seinfeld, J. H., Jayne, J. T., and Worsnop, D. R.: Aircraft-based aerosol size and composition measurements during ACE-Asia using an Aerodyne aerosol mass spectrometer, J. Geophys. Res., 108(D23), 8645, doi:10.1029/2002JD003226, 2003.

Bahreini, R., Keywood, M. D., Ng, N. L., Varutbangkul, V., Gao, S., Flagan, R. C., Seinfeld, J. H., Worsnop, D. R., and Jimenez, J. L.: Measurements of secondary organic aerosol

from oxidation of cycloalkenes, terpenes, and *m*-xylene using an Aerodyne aerosol mass spectrometer, Environ. Sci. Technol., 39(15), 5674–5688, doi:10.1021/es048061a, 2005.

Bates, T. S., Quinn, P. K., Coffman, D. J., Johnson, J. E., and Middlebrook, A. M.: Dominance of organic aerosols in the marine boundary layer over the Gulf of Maine during NEAQS 2002 and their role in aerosol light scattering, J. Geophys. Res., 110, D18202, doi:10.1029/2005JD005797, 2005.

Brown, S. S., Dibb, J. E., Stark, H., Aldener, M., Vozella, M., Whitlow, S., Williams, E. J., Lerner, B. M., Jakoubek, R., Middlebrook, A. M., DeGouw, J. A., Warneke, C., Goldan, P. D., Kuster, W. C, Angevine, W. M., Sueper, D. T., Quinn, P. K., Bates, T. S., Meagher, J. F., Fehsenfeld, F. C., and Ravishankara, A. R.: Nighttime removal of NO_x in the summer marine boundary layer, Geophys. Res. Lett., 31, L07108, doi:10.1029/2004GL019412, 2004.

Canagaratna, M. R., Jayne, J. T., Ghertner, D. A., Herndon, S., Shi, Q., Jimenez, J. L., Silva, P. J., Williams, P., Lanni, T., Drewnick, F., Demerjian, K. L., Kolb, C. E., and Worsnop, D. R.: Chase studies of particulate emissions from in-use New York City vehicles, Aerosol Sci. Technol., 38, 555–573, 2004.

de Gouw, J. A., Goldan, P. D., Warneke, C., Kuster, W. C., Roberts, J. M., Marchewka, M., Bertman, S. B., Pszenny, A. A. P., and Keene, W. C.: Validation of proton transfer reaction-mass spectrometry (PTR-MS) measurements of gas-phase organic compounds in the atmosphere during the New England Air Quality Study (NEAQS) in 2002, J. Geophys. Res., 108(D21), 4682, doi:10.1029/2003JD003863, 2003.

de Gouw, J. A., Middlebrook, A. M., Warneke, C., Goldan, P. D., Kuster, W. C., Roberts, J. M., Fehsenfeld, F. C., Worsnop, D. R., Canagaratna, M. R., Pszenny, A. A. P., Keene, W. C., Marchewka, M., Bertman, S. B., and Bates, T. S.: Budget of organic carbon in a polluted atmosphere: Results from the New England Air Quality Study in 2002, J. Geophys. Res., 110, D16305, doi:10.1029/2004JD005623, 2005.

Goldan P. D., Kuster, W. C., Williams, E., Murphy, P. C., Fehsenfeld, F. C., and Meagher, J.: Nonmethane hydrocarbon and oxy hydrocarbon measurements during the 2002 New England Air Quality Study, J. Geophys. Res., 109, D21309, doi:10.1029/2003JD004455, 2004.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys. Discuss., 6, 107–173, 2006.

Jayne, J. T., Leard, D. C., Zhang, X., Davidovits, P., Smith, K. A., Kolb, C. E., and Worsnop D. R.: Development of an aerosol mass spectrometer for size and composition analysis of

submicron particles, Aerosol Sci. Technol., 33, 49–70, 2000.
Jimenez, J. L., Jayne, J. T., Shi, Q., Kolb, C. E., Worsnop, D. R., Yourshaw, I., Seinfeld, J. H., Flagan, R. C., Zhang, X., Smith, K. A., Morris, J. W., and Davidovits, P.: Ambient aerosol sampling using the Aerodyne aerosol mass spectrometer, J. Geophys. Res., 108(D7), 8425, doi:10.1029/2001JD001213, 2003.
Klinedinst, D. B. and Currie, L. A.: Direct quantification of $PM_{2.5}$ fossil and biomass carbon within the Northern Front Range Air Quality Study's domain, Environ. Sci. Technol., 33, 4146–4154, 1999.
Kroll, J. H., Ng, N. L., Murphy, S. M., Flagan, R. C., and Seinfeld, J. H.: Secondary organic aerosol formation from isoprene photooxidation under high-NO_x conditions, Geophys. Res. Lett., 32, L18808, doi:10.1029/2005GL023637, 2005.
Lemire, K. R., Allen, D. T., Klouda, G. A., and Lewis, C. W.: Fine particulate matter source attribution for Southeast Texas using $^{14}C/^{13}C$ ratios, J. Geophys. Res., 107(D22), 4631, doi:10.1029/2002JD002339, 2002.
Lewis, C. W., Klouda, G. A., and Ellenson, W. D.: Radiocarbon measurement of the biogenic contribution to summertime PM-2.5 ambient aerosol in Nashville, TN, Atmos. Environ., 38, 6053–6061, doi:10.1016/j.atmosenv.2004.06.011, 2004.
Marcolli, C., Luo, B. P., Peter, Th., and Wienhold, F. G.: Internal mixing of the organic aerosol by gas phase diffusion of semivolatile organic compounds, Atmos. Chem. Phys., 4, 2593–2599, 2004.
McLafferty, F. W. and Tureek, F.: Interpretation of mass spectra, University Science Books, Sausalito, CA, 1993.
Murphy, D. M., Middlebrook, A. M., and Warshawsky, M.: Cluster analysis of data from the particle analysis by laser mass spectrometry (PALMS) instrument, Aerosol Sci. Technol., 37, 382–391, 2003.
Quinn, P. K. and Bates T. S.: North American, Asian, and Indian haze: Similar regional impacts on climate?, Geophys. Res. Lett., 30(11), 1555, doi:10.1029/2003GL016934, 2003.
Schneider, J., Borrmann, S., Wollny, A. G., Bläsner, M., Mihalopoulos, N., Oikonomou, K., Sciare, J., Teller, A., Levin, Z., and Worsnop, D. R.: Online mass spectrometric aerosol measurements during the MINOS campaign (Crete, August 2001), Atmos. Chem. Phys., 4, 65–80, 2004.
Shibata, K., Endo, M., Yamamoto, N., Yoshinaga, J., Yanagisawa, Y., Endo, O., Goto, S., Yoneda, M., Shibata, Y., and Morita, M.: Temporal variation of radiocarbon concentration

in airborne particulate matter in Tokyo, Radiocarbon, 46, 485–490, 2004.
Szidat, S., Jenk, T. M., Gäggeler, H. W., Synal, H.-A., Fisseha, R., Baltensperger, U., Kalberer, M., Samburova, V., Reimann, S., Kasper-Giebl, A., and Hajdas, I.: Radiocarbon (^{14}C)-deduced biogenic and anthropogenic contributions to organic carbon (OC) of urban aerosols from Zürich, Switzerland, Atmos. Environ., 38, 4035–4044, doi:10.1016/j.atmosenv.2004.03.066, 2004.
Williams, J., Roberts, J. M., Fehsenfeld, F. C., Bertman, S. B., Buhr, M. P., Goldan, P. D., Hübler, G., Kuster, W. C., Ryerson, T. B., Trainer, M., and Young, V.: Regional ozone from biogenic hydrocarbons deduced from airborne measurements of PAN, PPN, and MPAN, Geophys. Res. Lett., 24(9), 1099–1102, 1997.
Zhang, Q., Alfarra, M. R., Worsnop, D. R., Allan, J. D., Coe, H., Canagaratna, M. R., and Jimenez, J.-L.: Deconvolution and quantification of hydrocarbon-like and oxygenated organic aerosols based on aerosol mass spectrometry, Environ. Sci. Technol., 39, 4938–4952, doi:10.1021/es048568l, 2005a.
Zhang, Q., Canagaratna, M. R., Jayne, J. T., Worsnop, D. R., and Jimenez, J.-L.: Time- and size-resolved chemical composition of submicron particles in Pittsburgh: Implications for aerosol sources and processes, J. Geophys. Res., 110, D07S09, doi:10.1029/2004JD004649, 2005b.

Table 1. Assignment of fragments for the most intense m/z of the investigated mass spectra.

m/z	Fragments
26	CN^+, $CHCH^+$
27	CH_2CH^+
29	$CH_3CH_2^+$, CHO^+, CH_3N^+
30	NO^+, CH_2O^+, CH_4N^+
31	CH_3O^+
41	$CH_2CHCH_2^+$ and isomers
42	$CH_2CH_2CH_2^+$, CH_2CO^+
43	$CH_3CH_2CH_2^+$, CH_2CHO^+, CH_3CO^+
44	CO_2^+
55	$CH_2CHCH_2CH_2^+$,CH_2CHCO^+ and isomers
57	$CH_3CH_2CH_2CH_2^+$, $CH_2CHCH_2O^+$, $CH_3CH_2CO^+$ and isomers
67	$CH_2(CH)_3CH_2^+$ and others
69	CF_3^+ and others
119	$CF_3CF_2^+$ and others
169	$CF_3CF_2CF_2^+$ and others

ACPD
6, 4601–4641, 2006

Table 2. Criteria used for generating clusters.

Pass No.	Strict Criterion	Loose Criterion	No. Spectra for Loose Criterion	Resulting No. Categories
1	0.97	0.97	–	4120
2	0.97	0.97	–	2233
3	0.97	0.97	–	1715
4	0.93	0.77	50	48

Table 3. Characteristics of the 15 most abundant categories.

Category	No. spectra	Most intense m/z (intensities normalized to the largest peak)
1	9505	44 (100), 43 (58), 29 (44), 27 (38), 30 (29), 41 (25), 42 (22), 55 (16)
2	1052	29 (100), 43 (86), 44 (83), 27 (55), 30 (44), 41 (34), 42 (27), 55 (24)
3	513	43 (100), 44 (92), 27 (39), 30 (33), 41 (32), 42 (26), 26 (15), 55 (14)
4	414	43 (100), 44 (40), 27 (31), 30 (25), 29 (23), 42 (15), 41 (14), 55 (12)
5	275	41 (100), 43 (99), 44 (93), 27 (67), 30 (53), 29 (51), 42 (33), 55 (30)
6	259	44 (100), 27 (71), 43 (54), 30 (50), 55 (32), 42 (25), 26 (20), 53 (15)
7	163	30 (100), 44 (60), 43 (49), 27 (36), 29 (36), 41 (24), 55 (17), 42 (17)
8	137	27 (100), 43 (74), 44 (62), 30 (37), 29 (31), 42 (23), 41 (22), 26 (17)
9	82	29 (100), 44 (49), 27 (29), 30 (26), 43 (23), 42 (15), 41 (13), 26 (13)
10	60	44 (100), 69 (92), 43 (50), 27 (38), 29 (29), 30 (25), 41 (23), 169 (19)
11	53	43 (100), 27 (78), 30 (52), 55 (29), 31 (22), 42 (21), 41 (20), 26 (18)
12	38	55 (100), 44 (57), 43 (45), 29 (32), 27 (30), 30 (24), 42 (15), 41 (14)
13	25	44 (100), 30 (37), 43 (28), 53 (13), 55 (11), 27 (9), 31 (8), 67 (7)
14	24	169 (100), 44 (66), 43 (30), 69 (25), 29 (22), 27 (21), 30 (15), 41 (14)
15	15	57 (100), 44 (44), 43 (40), 30 (25), 29 (23), 27 (22), 41 (20), 42 (12)

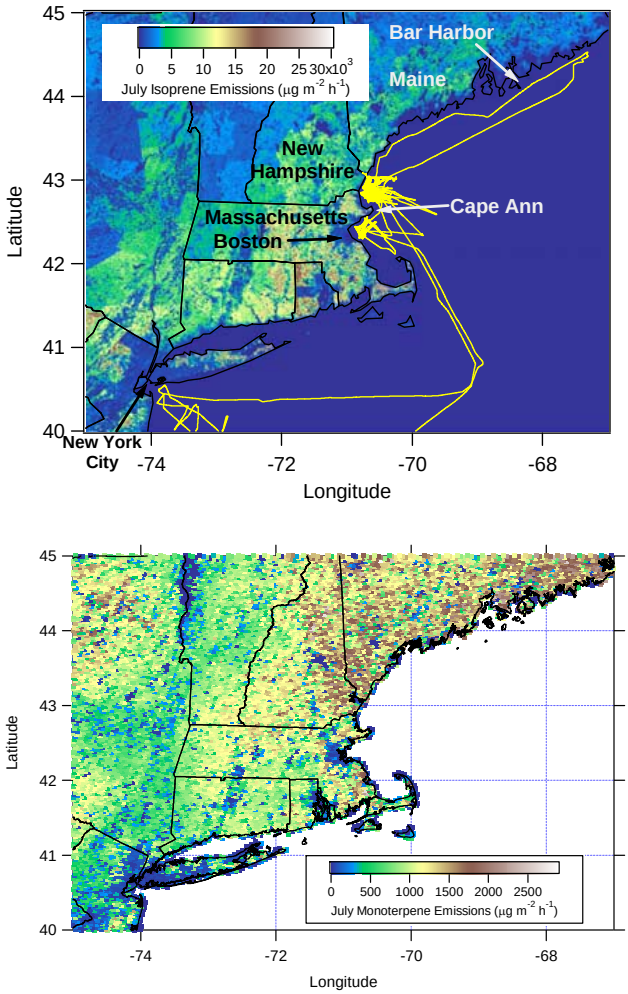

Fig. 1.

ACPD
6, 4601–4641, 2006

Cluster analysis of an AMS dataset

C. Marcolli et al.

EGU

Fig. 1. Maps showing the ship track (yellow) and isoprene emissions (top panel) and monoterpene emissions (bottom panel) for the month of July from the Megan biogenic source emission profile (Guenther et al., 2006). The dark blue patch at 71× longitude and 42.5× latitude is the Boston urban area. The ship sailed out of Charleston, SC on 12 July 2002 (off the map) to outside New York City on 16 July. On route to the study region, an issue with the AMS sampling inlet was discovered, rendering all of the data collected up to 18 July unusable. The ship sailed into the study region on 17 July, was downwind of Boston on 18 July, and then remained north of Cape Ann off the New Hampshire/Massachusetts coast until 24–26 July when it sailed up the coast of Maine past Bar Harbor. The ship was docked in New Hampshire from 26–29 July. The ship was positioned directly downwind of Boston on 31 July. From 1–5 August, the ship was off the New Hampshire coast and left the study region on 5 August. On 10 August, the ship sailed into Chesapeake Bay before heading back to Charleston and reaching the home port on 11 August (both cities are off the map).

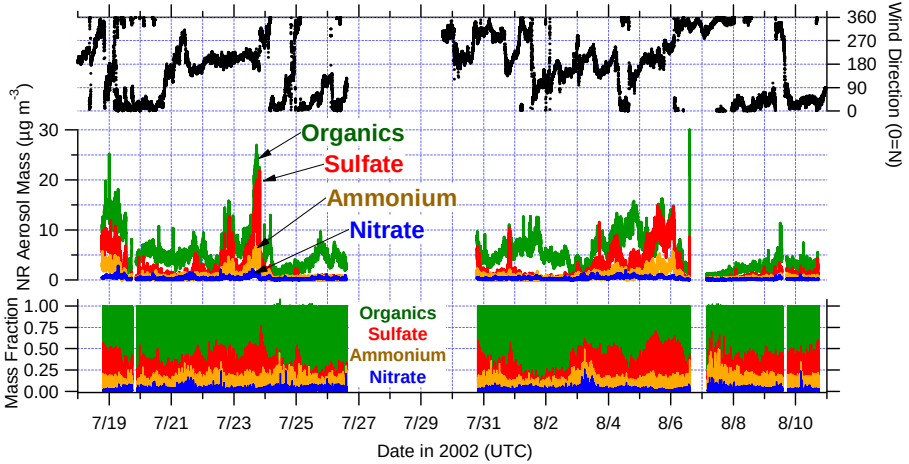

Fig. 2. Time series of the AMS non-refractory (NR) aerosol mass loadings, NR mass fraction, and wind direction.

ACPD
6, 4601–4641, 2006

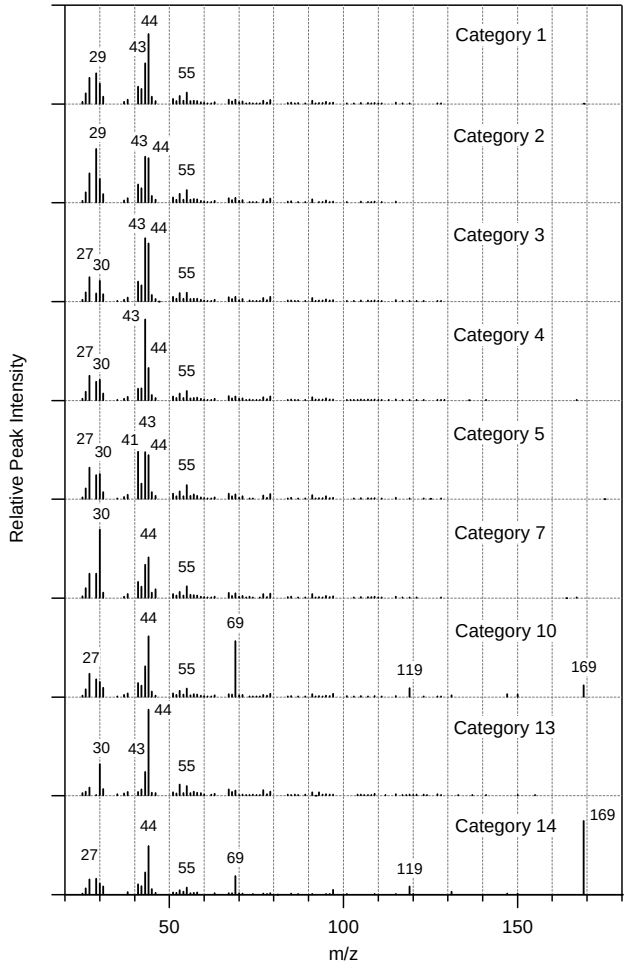

Fig. 3. Relative peak intensities (linear arbitrary units) for the top 5 and other prominent categories.

ACPD
6, 4601–4641, 2006

EGU

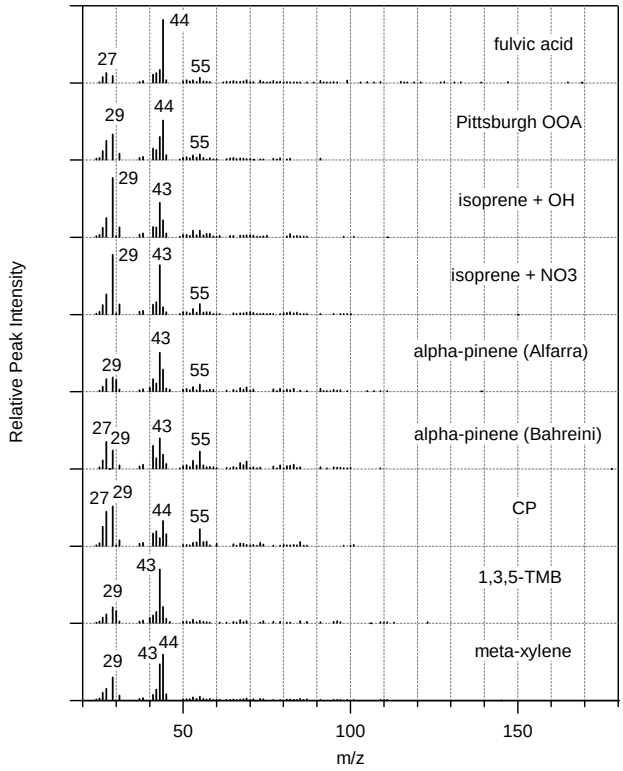

Fig. 4. Reference spectra (linear arbitrary units): laboratory generated spectra of fulvic acid (Alfarra, 2004b), Pittsburgh oxidized organic aerosol (OOA) from a field study (Zhang et al., 2005a), the aerosol oxidation products of isoprene plus hyrdroxyl radical (OH) (unpublished data), isoprene plus nitrate radical (NO3) (unpublished data), alpha-pinene (Alfarra et al., 2006[1]; Bahreini et al., 2005), cyclopentene (CP) (Bahreini et al., 2005), 1, 3, 5-trimethylbenzene (TMB) (Alfarra et al., 2006[1]), and meta-xylene (Bahreini et al., 2005).

ACPD
6, 4601–4641, 2006

EGU

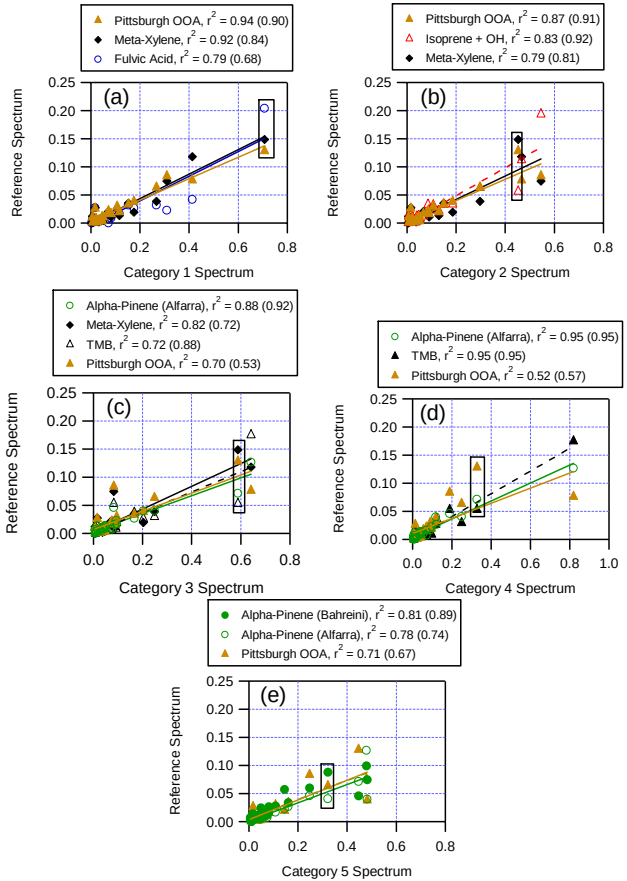

Fig. 5. Correlation plot of the peaks (arbitrary units) in the reference spectra (Fig. 4) with the peaks (arbitrary units) in categories 1–5 (Fig. 3). Note only the reference spectra with the highest correlation coefficients are shown. The correlation coefficients in parentheses are for correlations without m/z 44 (points in the box).

ACPD
6, 4601–4641, 2006

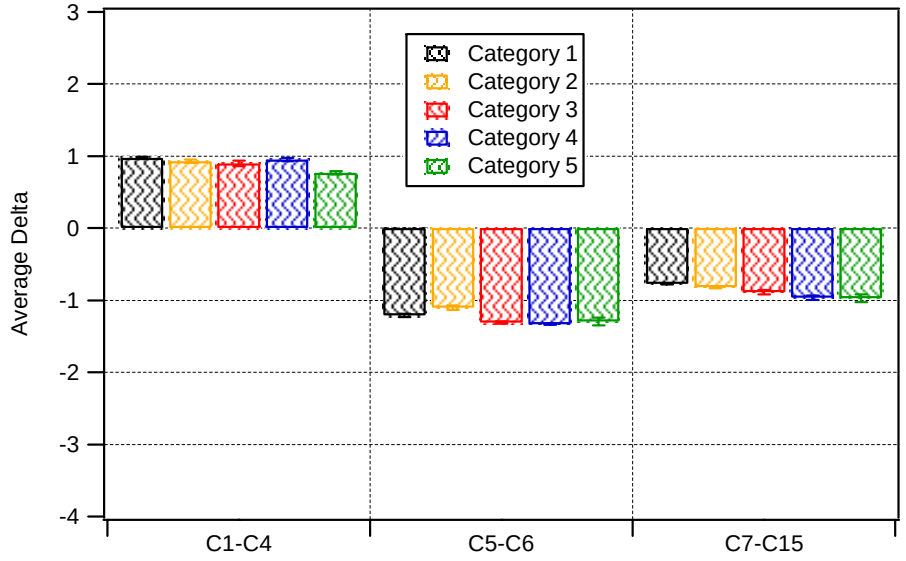

Fig. 6. Average delta patterns (arbitrary units) for low (C1–C4), mid (C5–C6), and high (C7–C15) m/z peaks for categories 1 through 5.

ACPD
6, 4601–4641, 2006

EGU

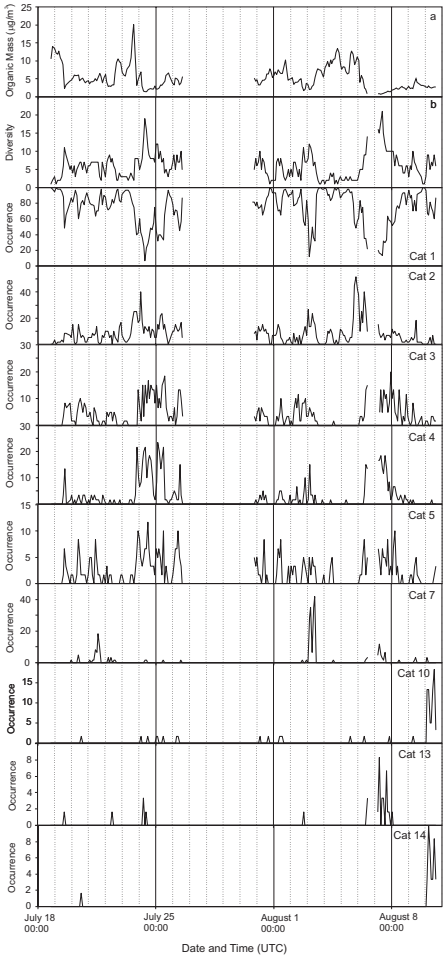

Fig. 7. Time series of the aerosol organic mass loading **(a)**, the diversity **(b)**, and the relative occurrence (%) of the top 5 and other prominent categories.

ACPD
6, 4601–4641, 2006

EGU

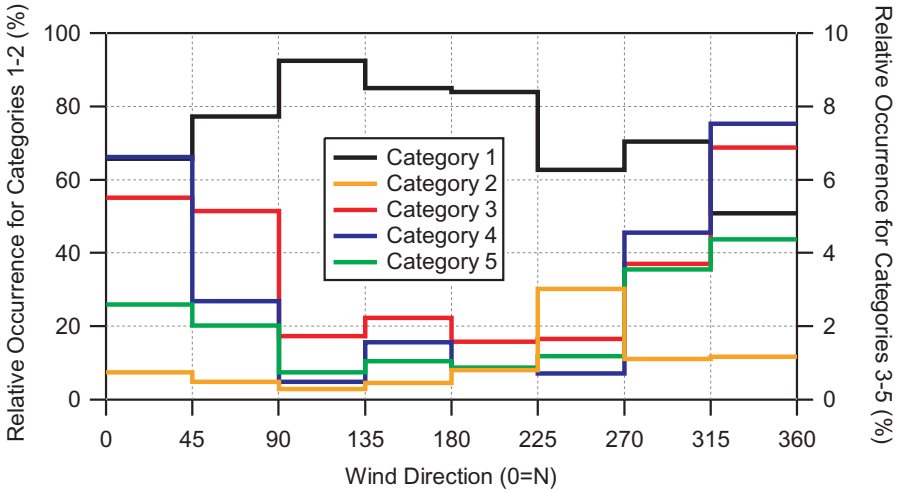

Fig. 8. Relative occurrence for categories 1 through 5 as a function of wind direction sorted for wind speeds greater than 4 m/s.

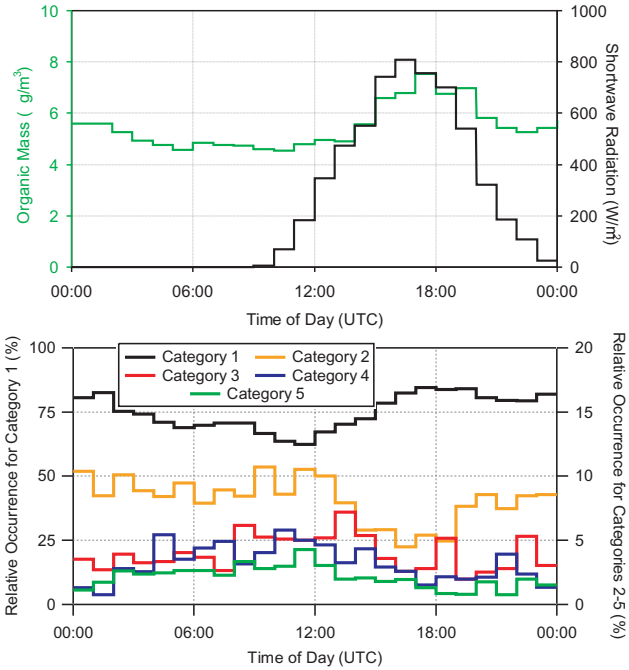

Fig. 9. Relative occurrence of particulate organic matter (top panel in green), and shortwave radiation (top panel in black) for a mean day averaged over the whole sampling period. Relative occurrence for categories 1 through 5 as a function of time of day (bottom panel).

ACPD
6, 4601–4641, 2006

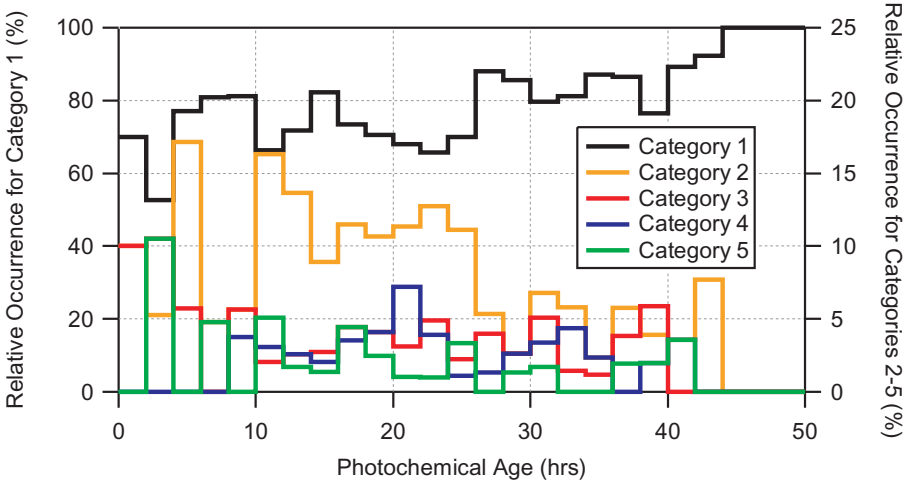

Fig. 10. Relative occurrence for categories 1 through 5 as a function of photochemical age.

ACPD
6, 4601–4641, 2006

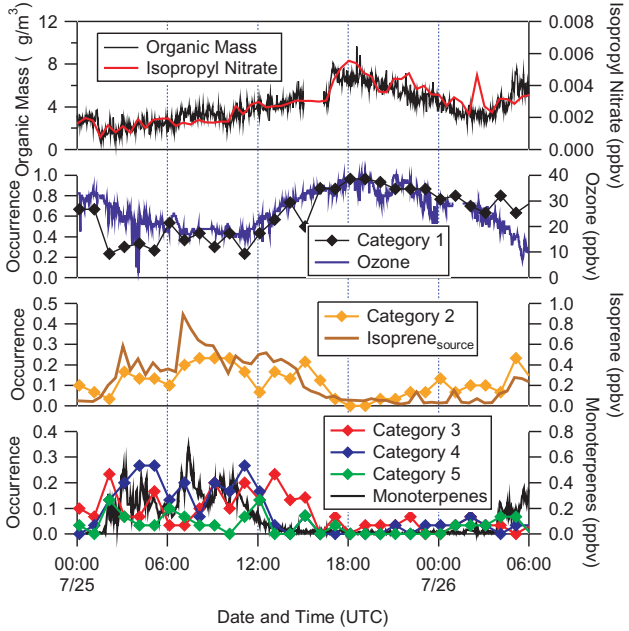

Fig. 11. Organic aerosol mass, isopropyl nitrate (a secondary anthropogenic volatile organic compound or VOC), ozone (from both biogenic and anthropogenic hydrocarbons), isoprene source (from de Gouw et al., 2005) and monoterpenes (biogenic VOCs), and the relative occurrence of categories 1 through 5 when the ship was sailing off the coast of Maine.

ACPD
6, 4601–4641, 2006

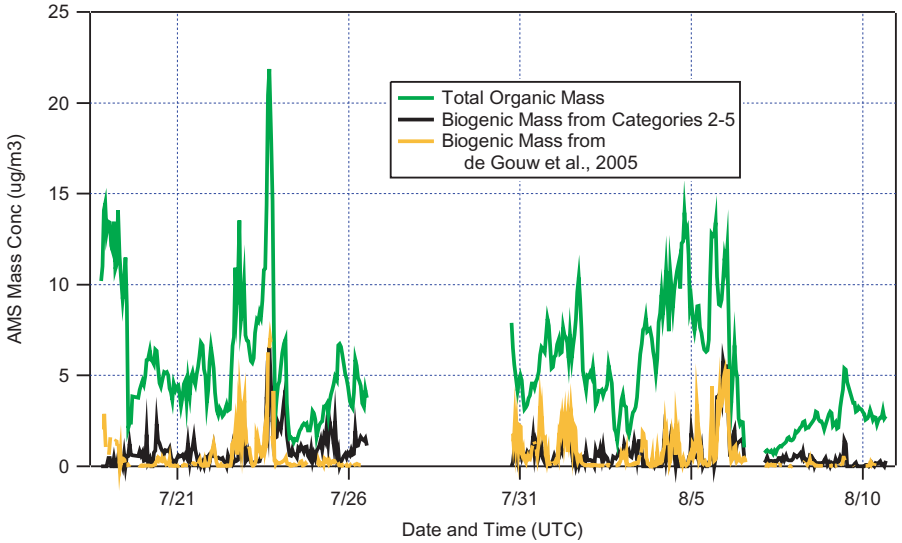

Fig. 12. Time trends of the total organic mass (green), the biogenic mass calculated from the mass in categories 2–5 (black), and the biogenic mass calculated from the biogenic fraction of the organic mass determined using the isoprene source concentrations by Gouw et al. (2005) (yellow).

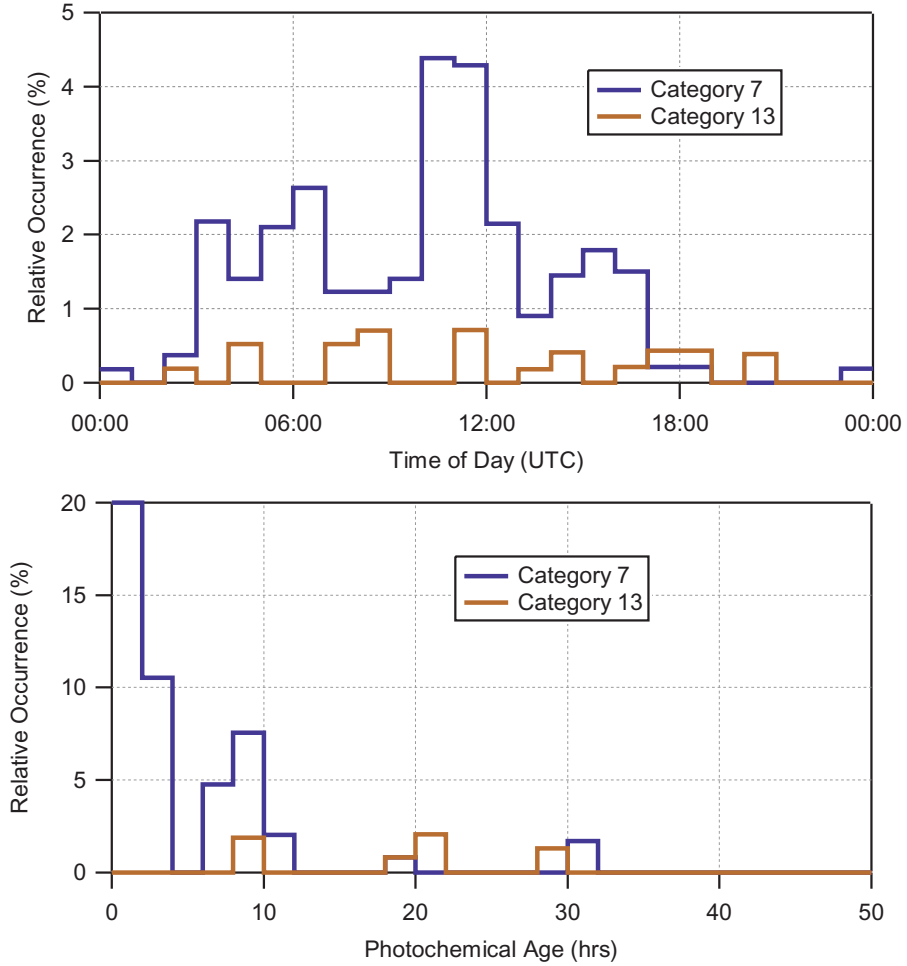

Fig. 13. Relative occurrence for categories 7 and 13 as a function of time of day (top panel) and photochemical age (bottom panel).